Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

Jun-Yan Zhu, et, al. ICCV 2017

2018.11.01

20185209 Sangyoon Lee

Table of contents

- Before presentation
- Review
- Relationship between Image Retrieval and CycleGAN
- **CycleGAN**
  - Introduction
  - Concept
  - Formulation
  - Network architecture
  - Result
  - Applications
  - Limitations
  - Summary

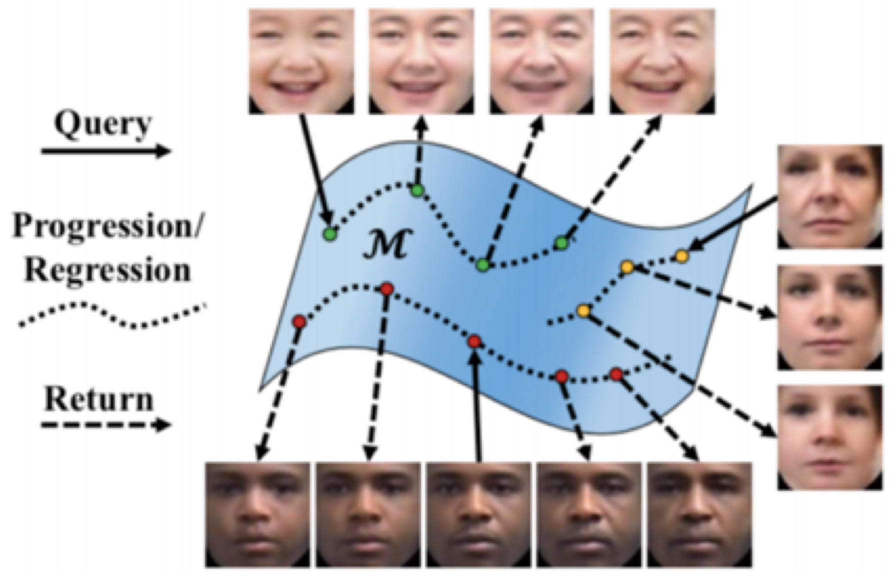
**ref: https://www.youtube.com/watch?v=Fkqf3dS9Cqw&t=1700s

Before presentation

- Original presentation topic
  - **GANerated Hands for Real-Time 3D Hand Tracking from Monocular RGB**
    - CVPR 2018
- However
  - This paper is dependent on CycleGAN.
- Therefore
  - Today's presentation topic
    - **Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks**
      - Jun-Yan Zhu, et, al. ICCV 2017

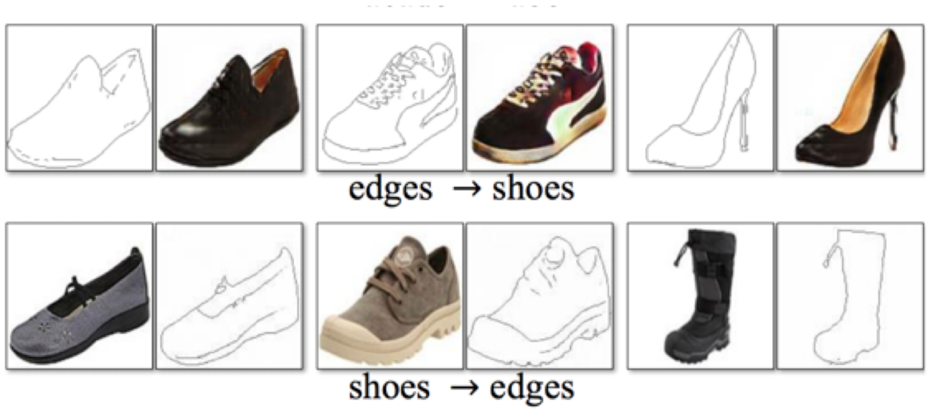
Review

- **Age Progression/Regression by Conditional Adversarial Autoencoder [CVPR `17]**
- Problems of Previous Works
  - Group-wised learning
  - Query with label
  - Step-by-step transition
- Solution
  - **Manifold Traversing**
    - The faces lies on a manifold
    - Traversing on the manifold corresponds to age/personality transformation



Relationship between Image Retrieval and CycleGAN

- Label annotation and paired data set are essential for effective network learning
- However, there is realistic limitations
- CycleGAN can be one of the examples to solve this problem
- There are various applications using CycleGAN for IR



Generated images by CycleGAN

Introduction

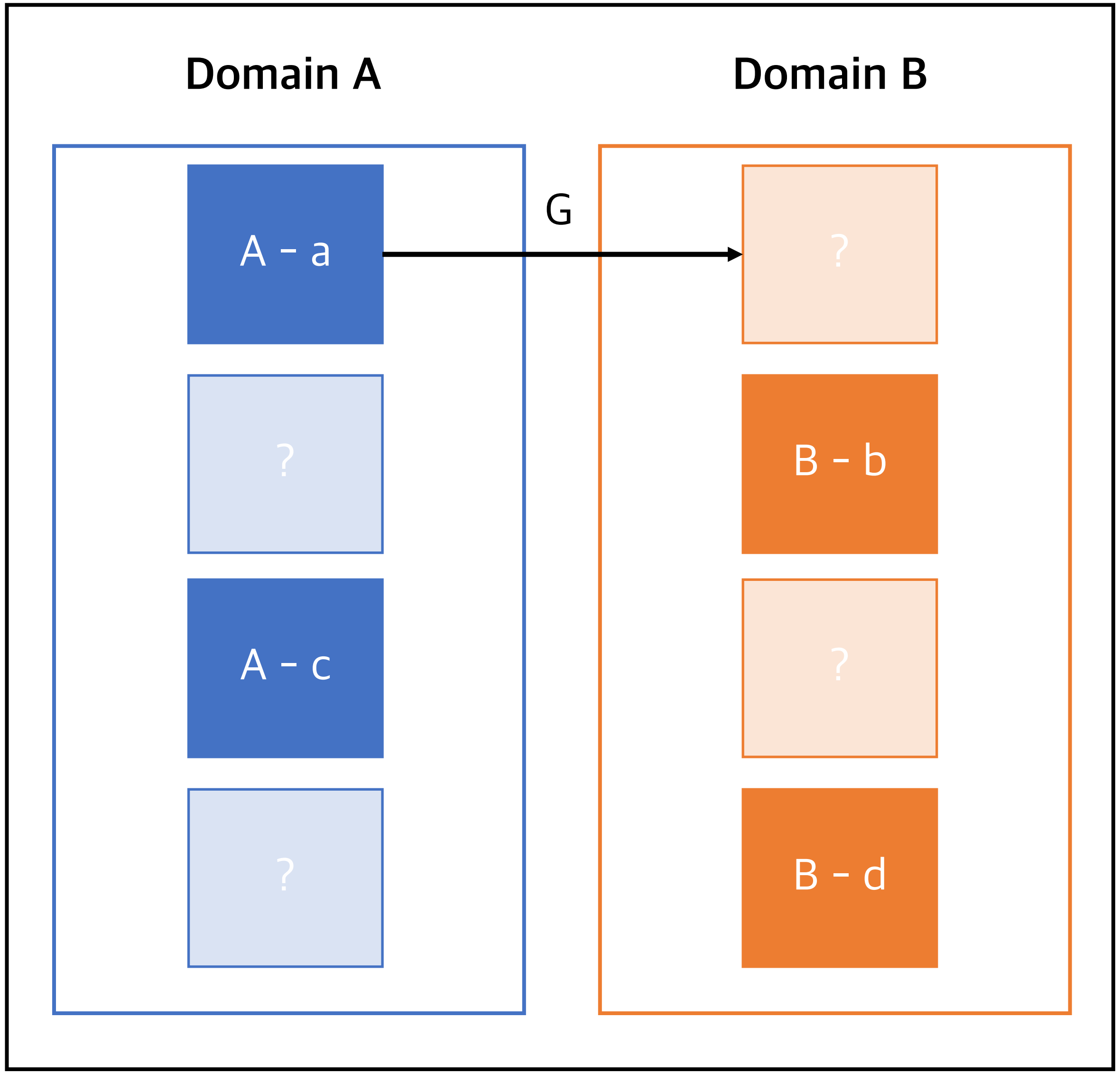
- CycleGAN
  - to learn how to translate domains from unpaired data sets
- Problem
  - Learning from an unpaired data set is important
    - it is very difficult to establish an exact matching set of paired data
  - Example
    - if you want to change a landscape image to Monet's style, you must have Monet's picture of the landscape you want.
- Solution
  - GAN
  - Cycle Consistency

Concept

Loss: $L_{GAN}(G(x), y)$

- G(x) should just look like a member in the Domain B

Unpaired Data Set

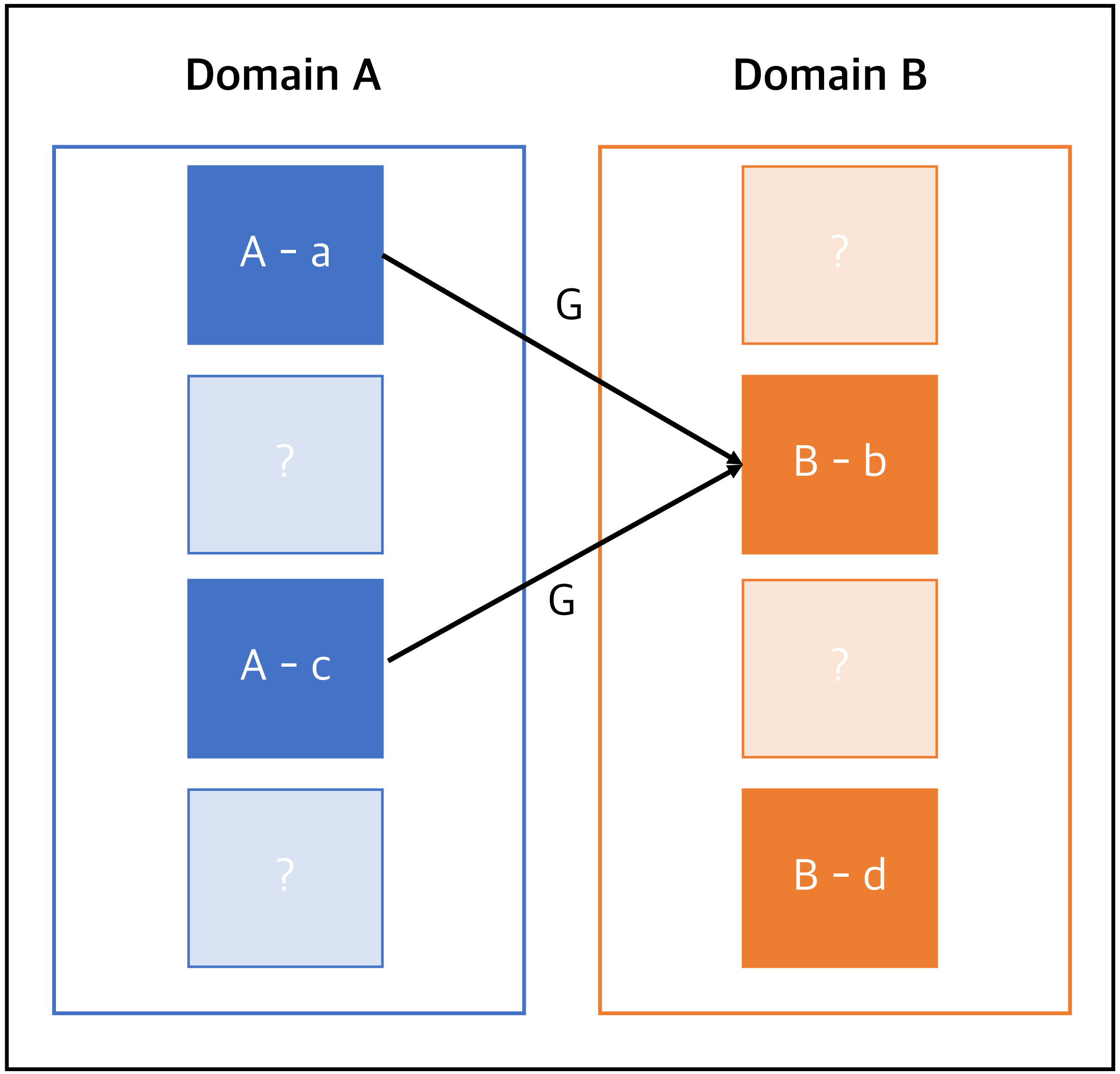


Concept

Loss: $L_{GAN}(G(x), y)$

- G(x) should just look like a member in the Domain B

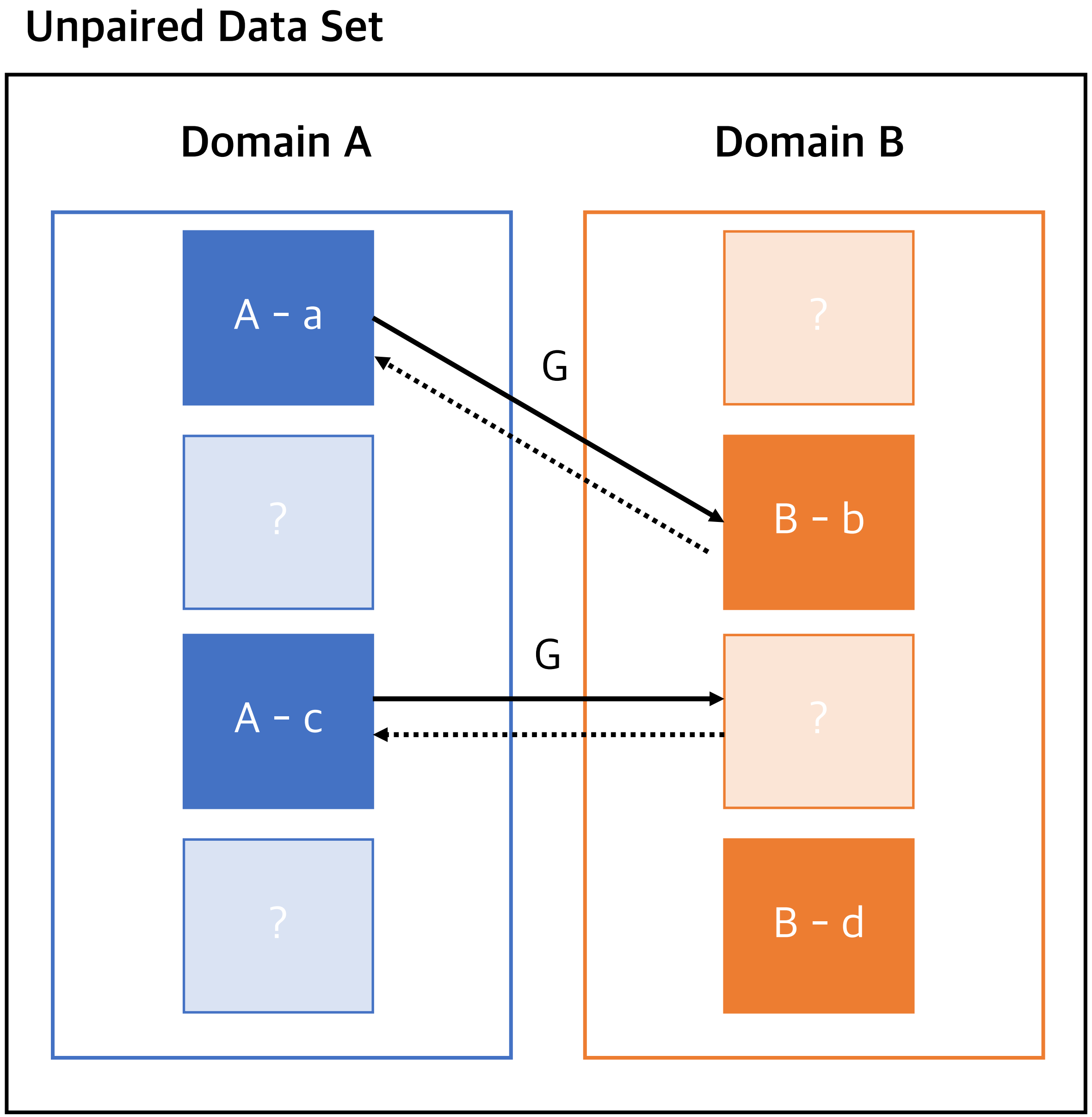
Unpaired Data Set



Concept

Loss: $L_{GAN}(G(x), y)$

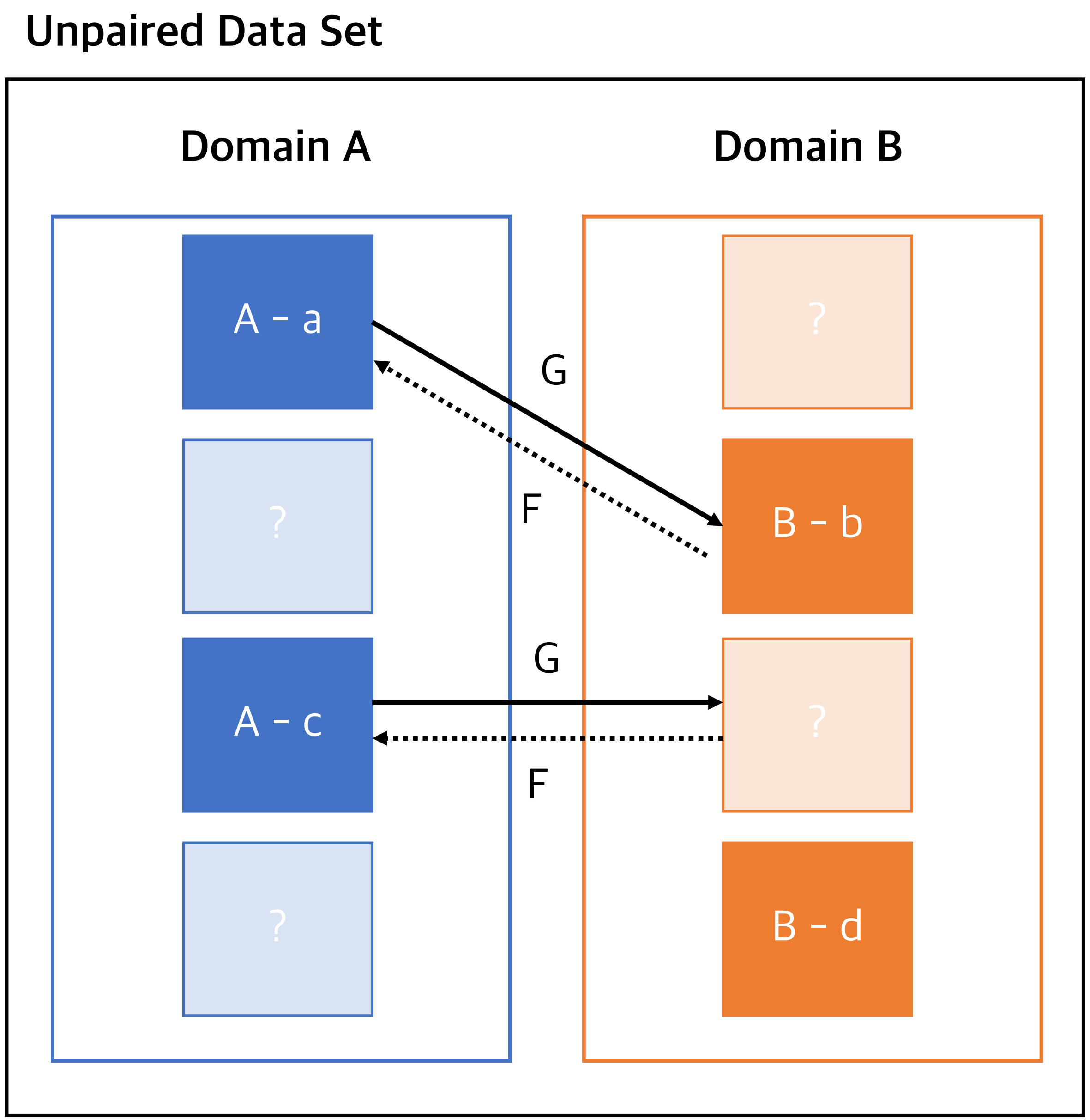
- G(x) should just look like a member in the Domain B
- And be able to reconstruct to original image in the Domain A



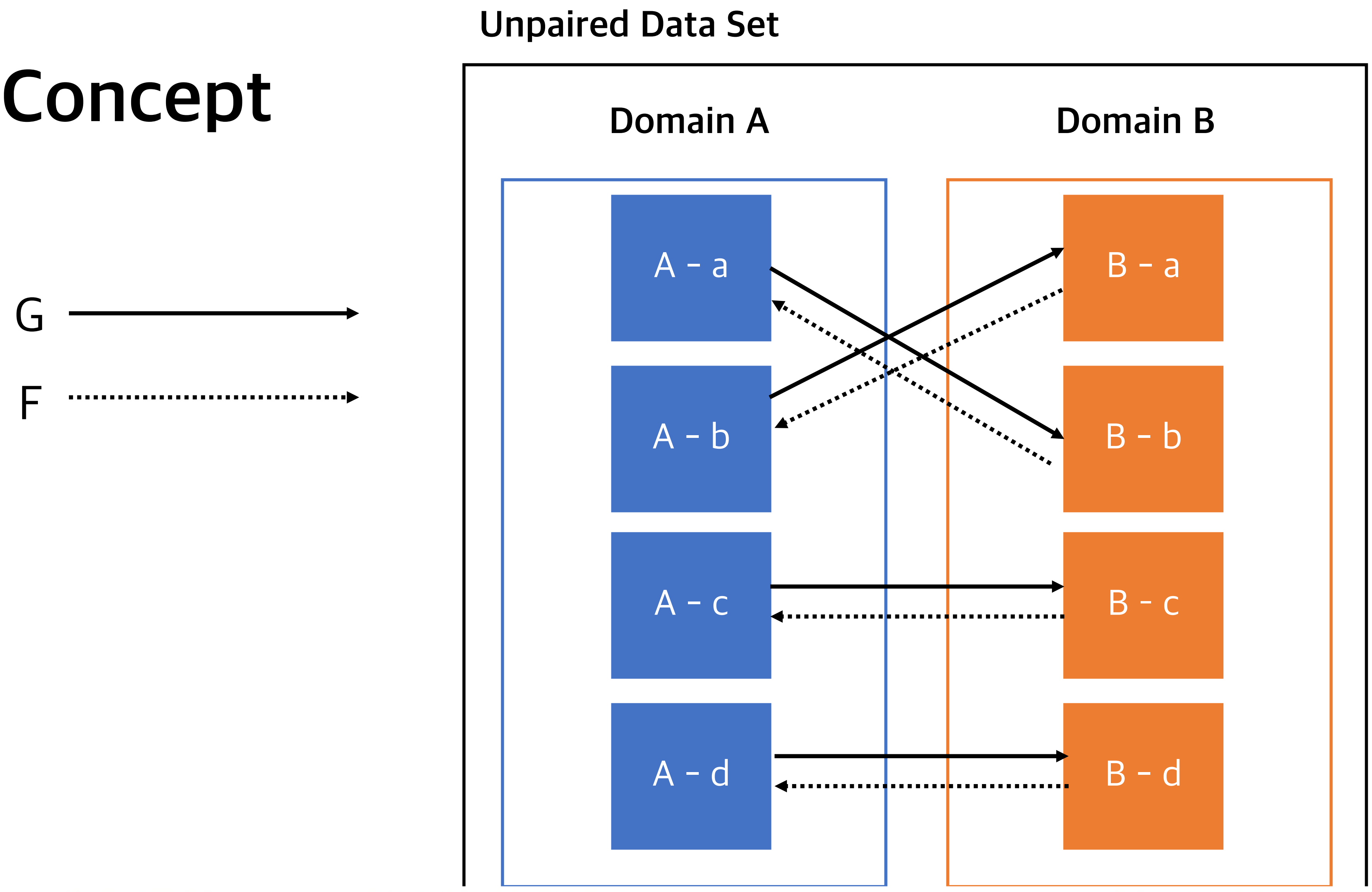
Concept

$$L_{GAN}(G(x), y) + \|F(G(x)) - x\|_1$$

- G(x) should just look like a member in the Domain B
- And be able to reconstruct to original image in the Domain A
- And F(G(x)) should be F(G(x)) = x, where F is the inverse deep network

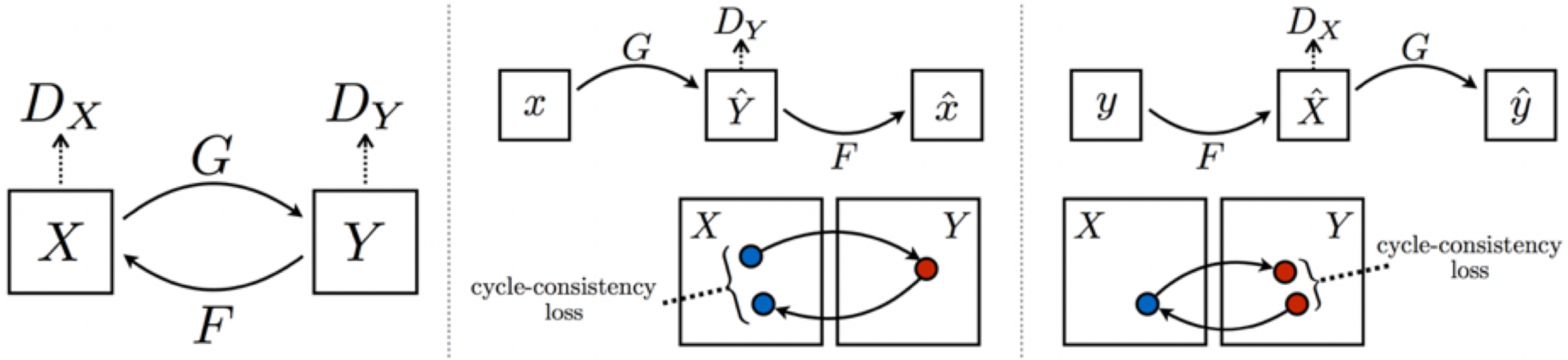


Concept

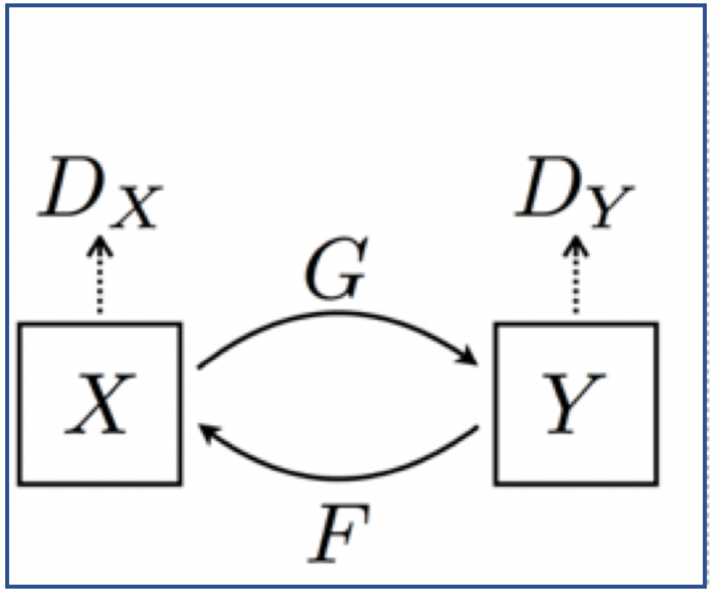


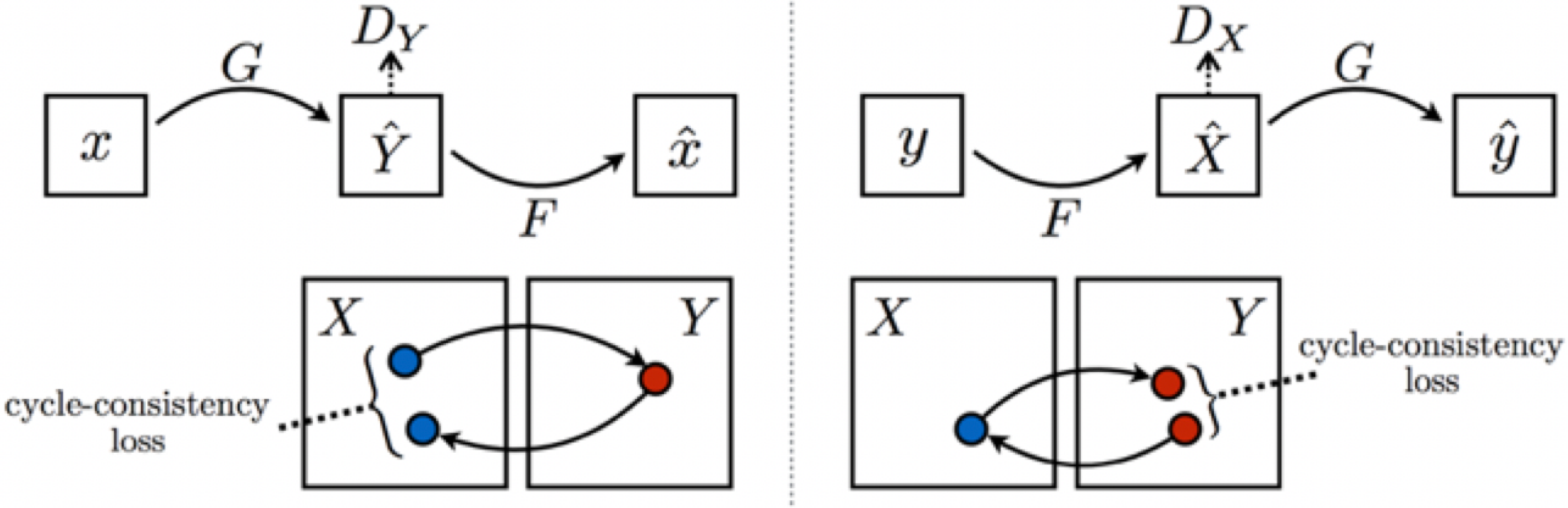
$$L_{GAN}(G(x),y)+\|F(G(x))-x\|_1 \quad + \quad L_{GAN}(F(y),x)+\|G(F(y))-y\|_1$$

Formulation - overview



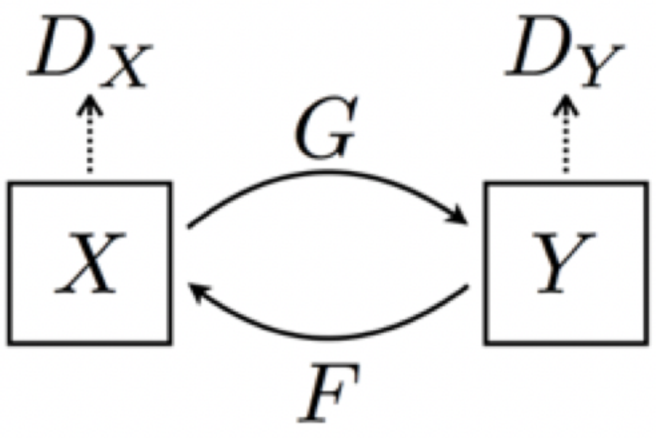
Formulation - Adversarial Loss

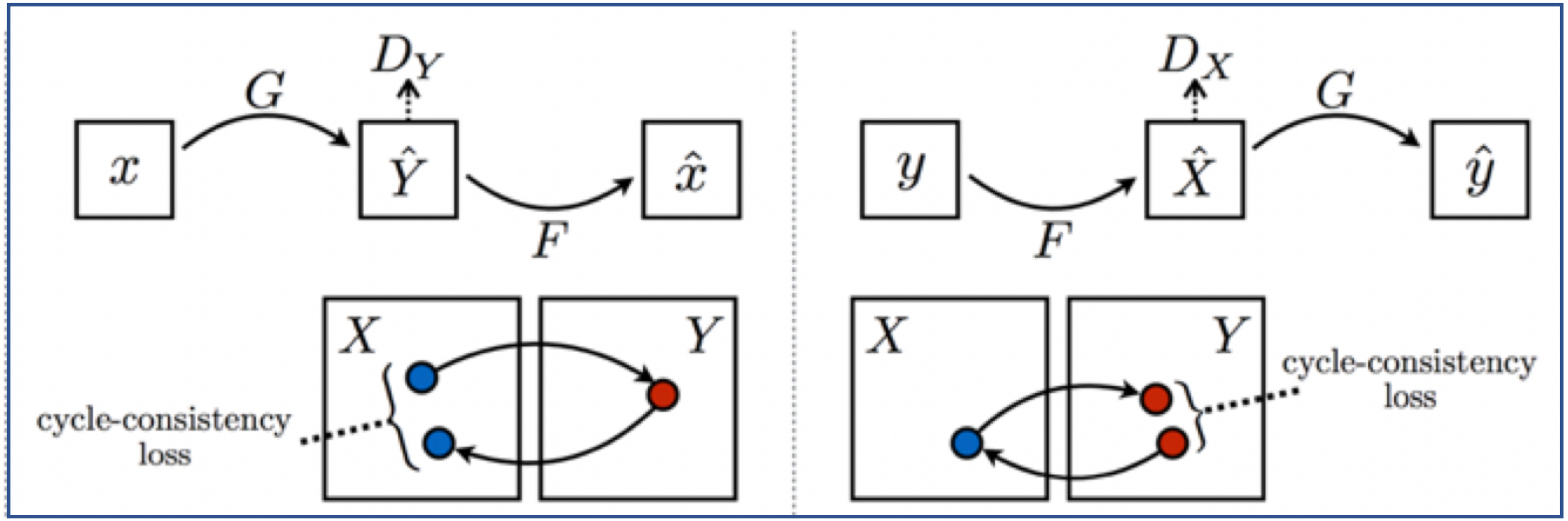




$$\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = \mathbb{E}_{y\sim p_{\text{data}}(y)}[\log D_Y(y)] + \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log(1 - D_Y(G(x))], \tag{1}$$

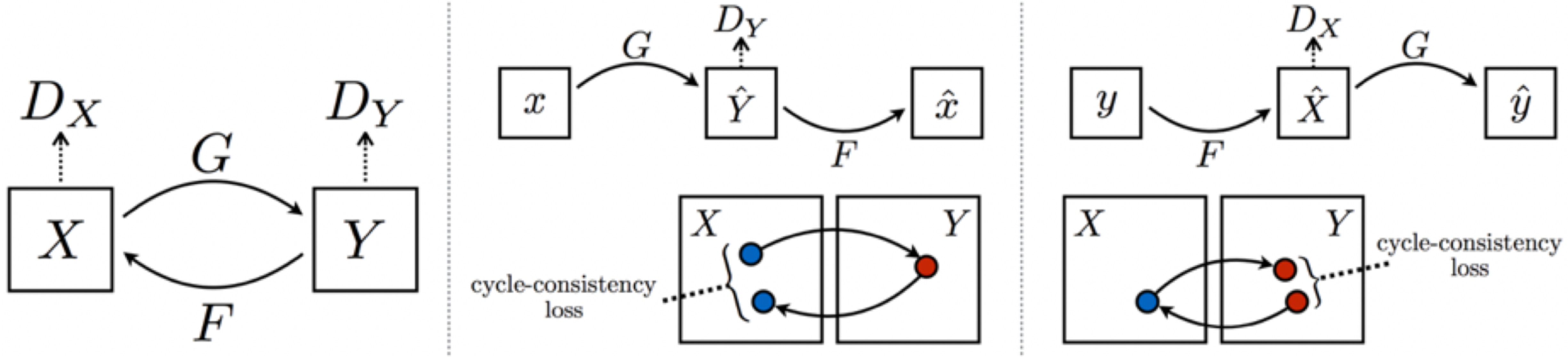
Formulation - Cycle Consistency Loss





$$\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1]. \quad (2)$$

Formulation - Full Objective

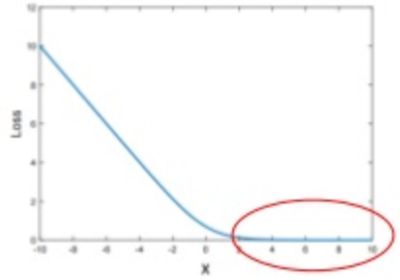


$$
\begin{aligned}
\mathcal{L}(G, F, D_X, D_Y) =& \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) \\
&+ \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) \\
&+ \lambda \mathcal{L}_{\text{cyc}}(G, F), \qquad (3)
\end{aligned}
$$

Network architecture

- ResNet for the generator
  - ResNet is effective for high resolution image processing
- PatchGAN (70 * 70) for the Discriminator
- Use Least Square GAN Loss instead cross entropy
  - With cross entropy

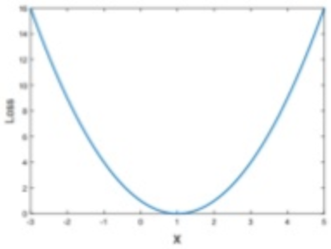
$$\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{\text{data}}(y)}[\log D_Y(y)] + \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log(1 - D_Y(G(x)))],$$



Vanishing gradients

  - With Least Square

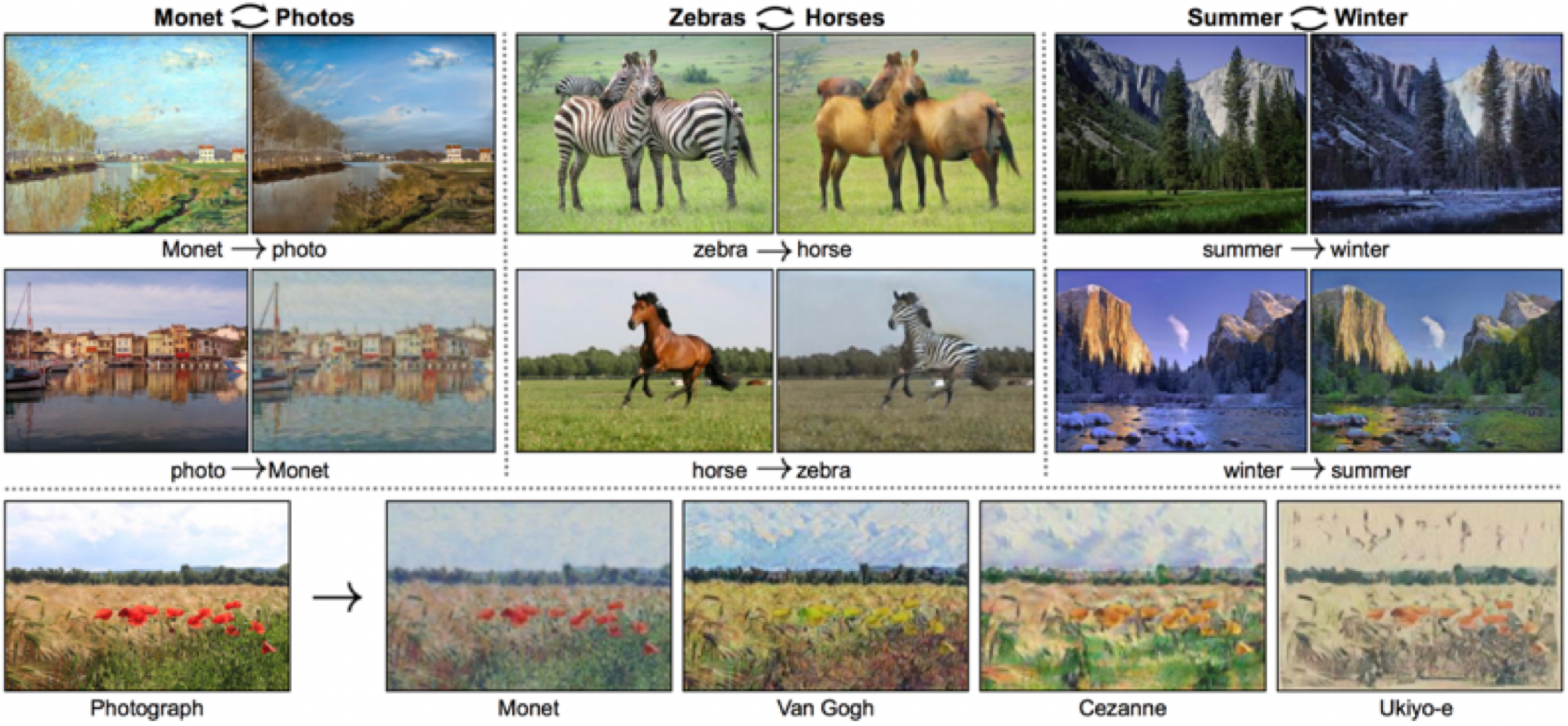
$$\mathcal{L}_{\text{LSGAN}}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{\text{data}}(y)}[(D_Y(y) - 1)^2] + \mathbb{E}_{x \sim p_{\text{data}}(x)}[D_Y(G(x))^2],$$



Result


Input
Cycle alone
GAN alone
GAN+forward
GAN+backward
CycleGAN
Ground truth

Result


Monet ⟷ Photos
Zebras ⟷ Horses
Summer ⟷ Winter
Monet → photo
zebra → horse
summer → winter
photo → Monet
horse → zebra
winter → summer
Photograph
→
Monet
Van Gogh
Cezanne
Ukiyo-e

Result

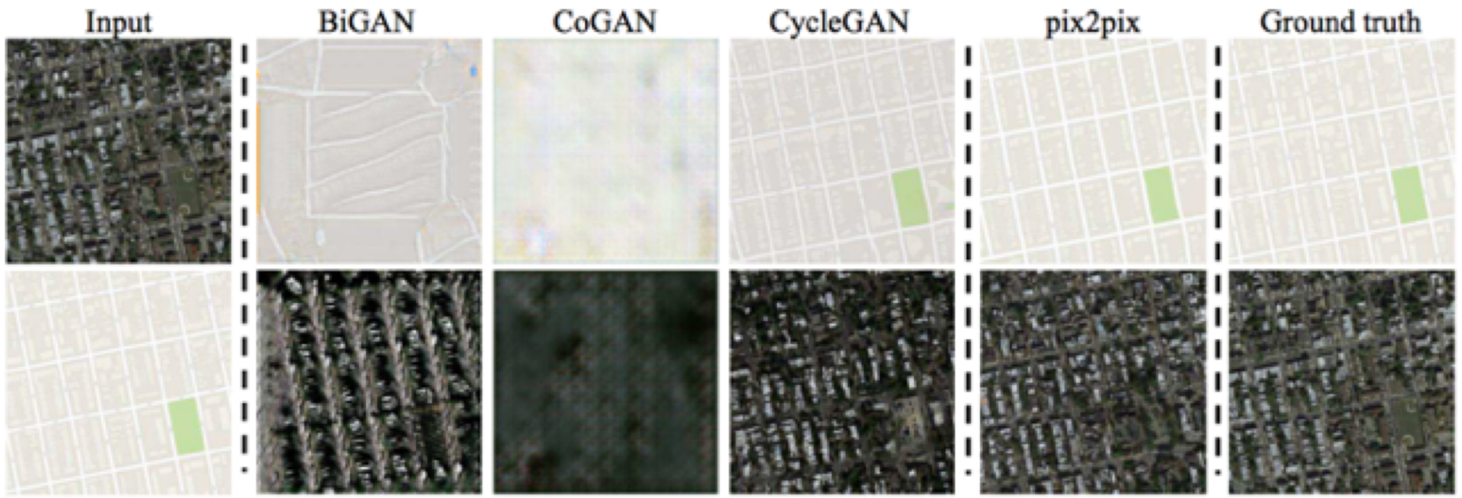
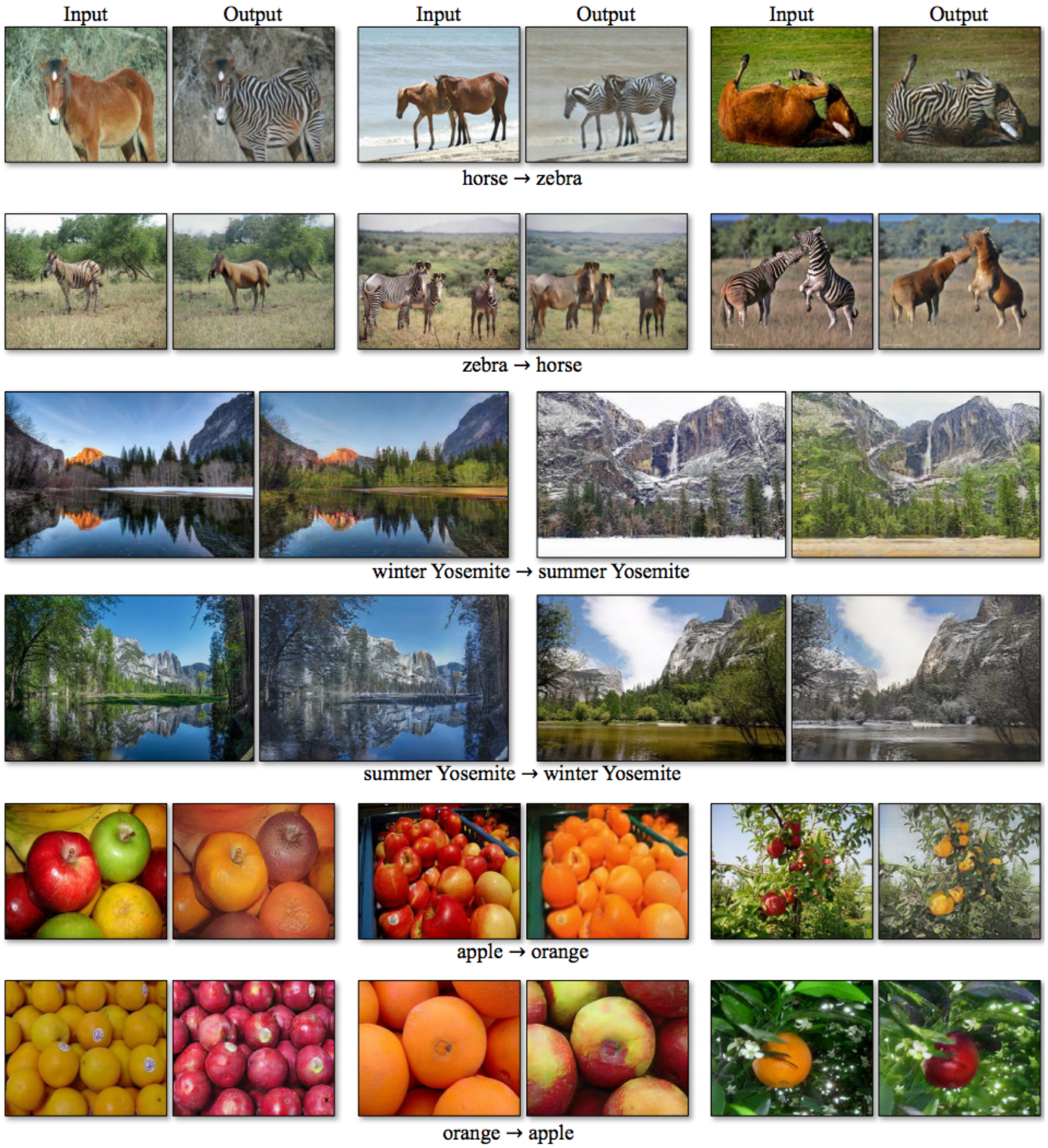


Figure 5: Different methods for mapping aerial photos↔maps on Google Maps. From left to right: input, BiGAN/ALI [6, 7], CoGAN [28], CycleGAN (ours), pix2pix [20] trained on paired data, and ground truth.

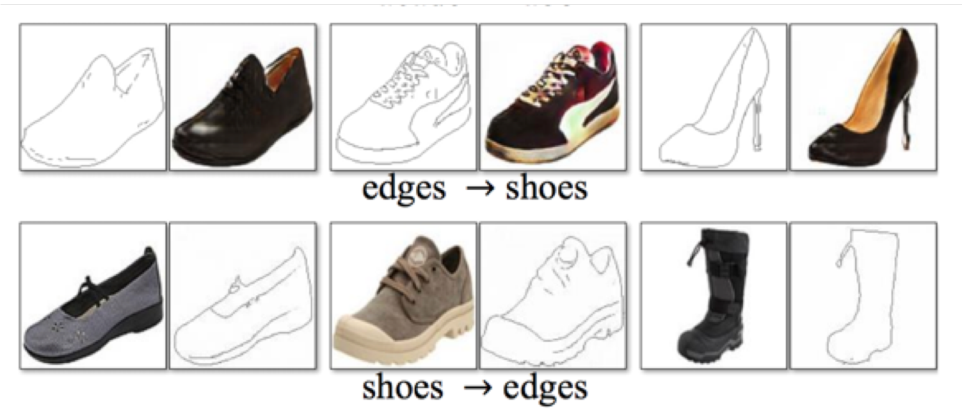
Loss	Map → Photo % Turkers labeled *real*	Photo → Map % Turkers labeled *real*
CoGAN [28]	0.6% ± 0.5%	0.9% ± 0.5%
BiGAN/ALI [7, 6]	2.1% ± 1.0%	1.9% ± 0.9%
Pixel loss + GAN [42]	0.7% ± 0.5%	2.6% ± 1.1%
Feature loss + GAN	1.2% ± 0.6%	0.3% ± 0.2%
CycleGAN (ours)	**26.8% ± 2.8%**	**23.2% ± 3.4%**

Table 1: AMT "real vs fake" test on maps↔aerial photos.

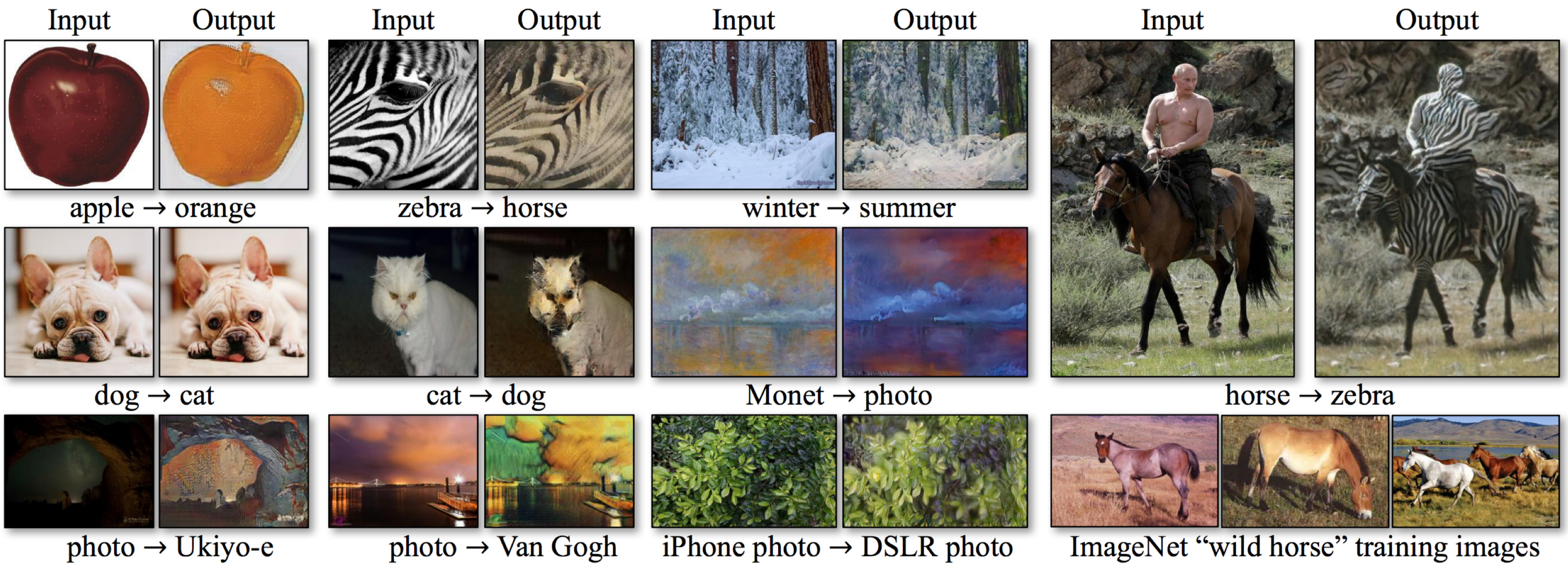
Applications



Applications



Limitations



- It is difficult to change the shape
- Sensitive to data distribution

Summary

- To incorporate Cycle Consistency into the existing GAN model and work with Unpaired Dataset.
- Use ResNet, LSGAN, PatchGAN for high resolution style transfer
- It is difficult to make a large change in shape due to constraints.
- Slow learning due to large network

Q & A

- Thank you for listening

Quiz

- Q1
  - **What is the newly proposed loss function for unpaired data set in this paper?**
    - A) Cycle Consistency
    - B) Rectangle Consistency
    - C) Triangle Consistency
    - D) Adversarial
- Q2
  - **Which of the following is not related to the disadvantages of CycleGAN?**
    - A) high resolution style transfer
    - B) Slow learning speed
    - C) it is difficult to change the shape
    - D) Sensitive to data distribution